## Project Abstract

**Creating Dynamic Social Network Models from Sensor Data**

(NSF Award # 0433637 and 0433376)
Tanzeem Choudhury, Henry Kautz, James Kitts and Dieter Fox
University of Washington and Intel Research Seattle
Jim Rehg
Georgia Institute of Technology

### Overall Mission/Objective

People are social animals. Interaction with others underlies many aspects of our lives: how we learn, how we work, how we play and how we affect the broader community. In this project, our aim is to develop a sensor-based approach for understanding people's interactions and their social networks: how people communicate with and influence each other, how they build networks of social relationships, and how such social networks develop over time.

We are asking these questions using a new and, we believe, powerful set of interdisciplinary tools, drawn from Sociology and Computer Science. Sociologists have developed a rich set of theoretical tools for studying social networks, such as methods for measuring people's prominence within social groups, and for quantifying people's tendency to interact with others who are similar. However, it has been difficult to collect data that are as rich as the theories. The only methods available have been either impractically laborious, such as directly observing people for hours on end, or potentially unreliable, such as asking people to answer questionnaires or to write diaries.

Methods from Computer Science can transform the data collection process, by making it possible to automatically record detailed aspects of people's social interactions, for hours and days on end. This is achieved by having the data collected not by people, but instead by computers. These computers are *wearable*, meaning that they are carried around on the bodies of the study participants: lightweight, small and unobtrusive, worn just as one might wear a watch, a necklace or a belt. These wearable computers incorporate sensors to record various aspects of the immediate environment, such as sound, movement, and geographical location. Wearable computers are already becoming part of our lives: iPods, cell phones, and runner's heart-monitoring watches are all forms of wearable computer, although they lack the wide variety of sensors built-in to the wearables in our study.

Wearable computers can collect data in several ways that humans cannot. First, they can record continuously with fine time-resolution, e.g. hundreds of recordings per second. In addition, they can record uninterruptedly, for long periods of time. Perhaps most importantly, a large number of people can wear such computers simultaneously, making it possible to study the social interactions not only between individuals but also within and between groups. We have already used such devices to record the social interactions between twenty-three people over a period of two weeks, and we are currently starting a study to investigate the formation and evolution of social networks in a large proportion of the incoming class of University of Washington

Computer Science graduate students, over the course of an academic year. These types of data sets will many times more richly detailed than hand-collected data about social networks, which, for face-to-face interactions, is all that has been available until now. Such datasets will enable us to investigate a host of important sociological questions. For example: how do social networks change over time, over short and long time scales? Do individuals sharing relations tend to be similar in attitudes, and behaviors? Do we choose friends who are similar to us, or do we become more similar to our friends? How does the physical environment affect the social networks of people who live and work there?

It is one thing to collect a mass of data, and quite another thing to understand what the data might be telling us. We are building computational models that can find structure in the masses of noisy sensor data that get collected. In particular, these models should be able to infer people's roles and relationships, and describe how individuals influence one another. Probabilistic models, e.g. Bayesian networks and Markov networks, are ideal for finding the underlying structure and dynamics of the social networks that led to the recorded person-to-person interactions. Because these computational models are probabilistic, they allow us not only to make sense of the raw sensor data, but also to handle the variations that are characteristic of human behavior.

**Progress and (Preliminary) Outcomes**:

Our current project at the University of Washington grew out of a pilot study of twenty-three graduate students, who wore wearable computers with sensors during business hours, over a period of two weeks. Computational and statistical modeling of the collected data yielded several interesting preliminary results. For example, it was found that people mostly stuck within their own groups, but that there were a few individuals who mixed between groups, thereby functioning as the community's social connectors. We also found that the people who were the most socially connected in the community were also those who tended to have the greatest influence on other people's styles of conversation. These findings, although promising, were limited by the relatively small scale and short duration of that study.

To look at larger-scale and longer-term trends, we are right now starting a new study with the incoming class of University of Washington graduate students at the Department of Computer Science. Thirty custom designed wearable computers have been constructed for the students to wear, with each device containing the following sensors: audio, WiFi (wireless internet), GPS (satellite global positioning system), accelerometers (to detect motion), temperature, light, humidity and air-pressure. At the present time (August, 2005) enrollment is in process, and data-collection will commence with the start of the 2005 Fall Quarter. The participants will be wearing the devices for one week per month, throughout the 2005-2006 academic year. Standard sociological surveys have also been developed and will be conducted periodically throughout the duration of the study. The survey data will give us preliminary measures of fixed characteristics (e.g. ethnicity, language, etc.) and variable characteristics (e.g. research interests), so that we can analyze how such characteristics and social networks co-evolve over time. We can then correlate the findings with behavioral measures obtained from sensor data. This study will create the first opportunity to study in detail how networks of social interactions develop in a large group, from the starting point of individuals who are meeting mostly for the first time, and then continuing over the course of a year during which social and work relationships are born and grow.

It is important to emphasize that we are taking great care to protect the personal privacy of the study participants. The collected audio is processed only to distinguish between periods of speaking and not-speaking, with no recording of intelligible speech or extraction of the words actually spoken. All recorded data will be anonymized and encrypted.

**Broader Impacts**

We expect that our study will have broad impact, both within and beyond the scientific community. As described above, our study will be able to address long-standing questions in Sociology at a far greater level of detail than was possible before, due to our automatically collected large-scale data set. Moreover, we are also developing statistical and computational models in order to understand the structure of the data.

We will be making our newly collected data sets available to the broader scientific community. Previous manually-collected data sets, although sparse and data-poor in comparison to what fully our automated methods make possible, have nonetheless been analyzed and re-analyzed by sociological theorists. We therefore anticipate that there will be substantial interest in the data-sets that we generate, and that new theoretical insights into social networks will emerge not only from our own computational analyses, but also from other researchers investigating our data.

The new understanding of social networks that, we believe, will emerge from our work is likely to have many potential applications in the real-world. Large businesses have long been interested in the flow of information within the organization, as the difference between their success and bankruptcy could depend on how well information flows between different groups of employees. Another real-world problem in which social networks play a central role is the spread of disease. An infectious outbreak in a self-contained village community would spread in a completely different way from an outbreak in a busy metropolitan city. Knowing the social networks in these communities can have enormous practical benefits. For example, the social network structure can help to predict the rate of spread of a disease, and also where it will spread to next.

These examples are just two of the possible applications of social network research. Many other domains could also be impacted, covering areas as diverse as marketing campaigns and urban planning. The wealth of new data made accessible by our new methods will, we expect, open up many more new questions than it will answer. We hope that it will also inspire further collaboration between the disciplines of Sociology and Computer Science. Our work will, perhaps, help to serve as a connector between the social networks of those two fields.

**Project Website**

http://www.cs.washington.edu/ai/socialnetworks/